LEVEL

GOAL-DIRECTED DECISION STRUCTURING SYSTEMS

Judea Pearl
Jin Kim
Robert Fiske

AD A103809

Technical Report covering the period 5/1/78 to 8/31/81

Work performed at Cognitive Systems Laboratory

School of Engineering and Applied Science

University of California, Los Angeles

Professor Judea Pearl, Principal Investigator

DTIC
SELECTED
SEP 2 1981

H

DTIC FILE COPY

81 9   02 015

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER	2. GOVT ACCESSION NO. AD-A103 809 / 3. RECIPIENT'S CATALOG NUMBER

4. TITLE (and Subtitle)	5. TYPE OF REPORT & PERIOD COVERED
GOAL-DIRECTED DECISION STRUCTURING SYSTEMS	Technical Report 5/1/78 - 6/30/81
	6. PERFORMING ORG. REPORT NUMBER UCLA-ENG-81-21

7. AUTHOR(s)	8. CONTRACT OR GRANT NUMBER(s)
Judea Pearl · Jin Kim · Robert Fiske	N00014-78-C-0372

9. PERFORMING ORGANIZATION NAME AND ADDRESS	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
University of California, Los Angeles School of Engineering and Applied Science Los Angeles, California 90024	NR 197-049

11. CONTROLLING OFFICE NAME AND ADDRESS	12. REPORT DATE
Office of Naval Research 800 N. Quincy Street Arlington, Virginia 22217	August 1981
	13. NUMBER OF PAGES

14. MONITORING AGENCY NAME & ADDRESS(If different from Controlling Office)	15. SECURITY CLASS. (of this report)
	UNCLASSIFIED
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for Public Release; Distribution Unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Decision Analysis, Decision Support Systems, Problem Structuring
Means-Ends Analysis, Knowledge-Based Systems, Alternatives Generation

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)   The report summarizes the development and evaluation of computerized decision structuring systems based on a new representational structure which offers several advantages over the traditional decision-tree representation. The design and operating characteristics of GODDESS (A Goal-Directed Decision Structuring System) and several environment simulators used for evaluating decision aiding tools are briefly outlined. The main body of the report focuses on an experimental evaluation of the effectiveness of two structuring procedures: 1) decision-tree elicitation and 2) goal-directed structuring. The goal-directed procedure appeared superior in encouraging subjects

DD FORM 1473   EDITION OF 1 NOV 65 IS OBSOLETE
1 JAN 73
S/N 0102-LF-014-6601

Abstract (cont.)

to generate novel (non-habitual) sets of effective options. The tree-elicitation procedure, on the other hand, permitted subjects to articulate more valid judgments and assessments, which in turn facilitated a more accurate recognition of the best action among the options given. The combined use of goal-directed procedures for structuring problems and tree-elicitation for optimization promises to utilize the strengths of both methods.

Accession For

NTIS GRA&I	✓
DTIC TAB	☐
Unannounced	☐
Justification	

By

Distribution/

Availability Codes

Dist	Avail and/or Special
A	

GOAL-DIRECTED DECISION STRUCTURING SYSTEMS

by

Judea Pearl

Robert Fiske

Jin Kim

Abstract

The report summarizes the development and evaluation of com-
puterized decision structuring systems based on a new representational
structure which offers several advantages over the traditional decision-
tree representation. The design and operating characteristics of GODDESS
(A Goal-Directed Decision Structuring System) and several environment
simulators used for evaluating decision aiding tools are briefly outlined.
The main body of the report focuses on an experimental evaluation of the
effectiveness of two structuring procedures: 1) decision-tree elicitation
and 2) goal-directed structuring. The goal-directed procedure appeared
superior in encouraging subjects to generate novel (non-habitual) sets
of effective options. The tree-elicitation procedure, on the other
hand, permitted subjects to articulate more valid judgments and assess-
ments, which in turn facilitated a more accurate recognition of the
best action among the options given. The combined use of goal-directed
procedures for structuring problems and tree-elicitation for optimiza-
tion promises to utilize the strengths of both methods.

TABLE OF CONTENTS

1.0 INTRODUCTION

This report summarizes the work performed toward the development and evaluation of Goal-Directed Decision-Structuring Systems during the period 5/1/78 to 6/30/81. The project was conducted under research contract N00014-78-C-0372 funded by the Engineering Psychology Programs Division of the Office of Naval Research. The research was performed at the Cognitive Systems Laboratory, University of California, Los Angeles, with Professor Judea Pearl as Principal Investigator.

The ultimate objective of this project has been to develop and evaluate a computerized decision structuring system based on a new and more effective representational structure which promises to offer several advantages over the traditional decision-tree representation. Our research followed two parallel avenues: 1) The development of computerized decision-structuring systems and tools for their evaluations, and 2) An experimental evaluation of the performance of the Goal-Directed Structuring approach. Since most of the development work is already documented in other reports, the focus of this report will be the evaluation phase. Section 2 will briefly summarize the highlights of the developmental works with references to the appropriate documentations. Section 3 will describe in detail the evaluation experiments conducted in the past six months and will state our conclusions.

2.0 SUMMARY OF DEVELOPMENTAL WORKS

This section contains brief highlights of three developmental works:

2.1 GODDESS: A Goal-Directed Decision Structuring System

2.2 Environment Simulators for Evaluation Studies

2.3 Experiments in Judgment Validity

The details of these developments are documented in other papers and reports, which are listed chronologically in Section 2.4. Reference numbers cited in Section 2.1, 2.2 and 2.3 refer to the list in Section 2.4. Section 2.5 contains a list of the research staff contributing to this project.

2.1 GODDESS: A Goal-Directed Decision Structuring System

GODDESS is an operational version of a computerized, domain-independent, decision structuring system based on a novel, goal-directed structure for representing decision problems. The structure allows the user to state relations among aspects, effects, conditions, and goals in addition to actions and states, which are the basic components of the traditional decision-tree approach. The program interacts with the user in a stylized English-like dialogue, starting with the stated objectives and proceeding to unravel the more detailed means by which these objectives can be realized. At any point in time, the program focuses the user's attention on the issues which are most crucial to the problem at hand.

The motivation for breaking away from the confines of decision-tree representations is elaborated in Ref. 3. It is based on the fact that the goal-directed structure is more refined and more compatible with the way people encode knowledge about problems and actions, and thus, enables the user to express judgments and beliefs which more closely represent the user's experience. Moreover, since action alternatives are evoked by first explicating the user's goals and intentions, the user may be guided toward the discovery of action alternatives he otherwise would not have identified.

2

The design principles of GODDESS, including its value-propagation procedures and dialogue management methodology,and a sample consultation are contained in Ref. 3. A more detailed account of GODDESS'latest implementation is provided by Ref. 12 "GODDESS Program Guide and User Manual", which is meant as a guide to those users who are considering implementing the system in their own computing environment. It also gives precise instructions for using GODDESS and highlights the options available to the user including modifications of query phrasings.

2.2 Environment Simulators for Evaluation Studies

Our approach for evaluating the merit of decision-aiding tools requires the development of computer-based systems for simulating a hypothetical decision-making environment. The essential features underlying this method are that operational tests are performed in an environment which is tightly controlled and thoroughly known to the evaluator, and that the merit of any decision plan enacted by the player is measured by an indisputable and computable 'ground-truth' performance criterion.

The subjects will first be trained to play the game and gain familiarity with the environment in which they operate. The training session is terminated when the performance score becomes constant over a significant length of time. At this point the Decision Support System will be turned on, and changes in the subject's performance will be monitored. The incremental improvement in the subject's score will provide one measure of the operational merit of the decision-aiding technique under study.

For the purpose of measuring the qualities of various decision strategies, we have built two simulated business games,whereby the player is instructed to accomplish certain objectives with limited resources and in-complete information. Several games in business environments have reached both a high level of popularity and a respectable status as faithful representations

3

of realistic decision-making environments.

Our major design goals for the first simulator were:

1. <u>Realism</u> - to make the game more challenging and to allow the player to exploit prior knowledge.

2. <u>Real-time Response</u> - to speed up the player's learning period.

In order to meet these two goals a sopisticated business game was developed. It is an adaptation of a popular game called "The Executive Game" by Henshaw and Jackson (Richard D. Irwin, Inc. 1979) which requires the player to adjust eight decision variables with each move. In addition, it provides an elaborate report on the state of the firm at the end of each game period. The simulator is described in full detail in Ref. 4, "A Graphic System for Evaluating Decision Aids."

Although the use of graphic interface was effective in shortening the learning-time, this system was still too advanced for our purposes. The major shortcoming was its complexity, which prevented us from computing an optimal game-playing strategy and hence made it impractical for us to assign to each decision an objective figure of merit.

Our second simulator constitutes a compromise between the requirements of realism and simplicity. We limited the player's actions to only four decision variables and designed an artificial model of the competing firm to allow the computation of an optimal game-playing strategy. The availability of an optimal strategy allows us to assign to each state of the game an objective figure of merit simply by turning the game over to a "super businessman" who plays the optimal strategy and recording his accumulated score. The merit of every action, therefore, can now be measured by reference to this optimal score. The difference between the accumulated score achievable by the optimal strategy and that achievable from the state created by any given action is defined as the loss-of-opportunity (LOO) associated with that action.

4

The mathematics behind this business simulator and its operating characteristics are documented in Ref. 11. This system was finally used in the evaluation experiments described in Section 3.

2.3 Experiments in Judgment Validity

The set of experiments described in this section were designed to answer some basic questions regarding the rationale for decision-aiding systems. Most decision-support technologies are founded on the paradigm that direct judgments are less reliable and less valid than synthetic inferences produced from more "fragmentary" judgments. Therefore, the reliability of the systems inferences should be highly sensitive to the reliability of their constituent rules. The latter may vary with the mode of reasoning invoked during the elicitation process, i.e., with the format in which the queries are phrased.

The first set of experiments (see Ref. 5: "Experiments in Cognitive Decomposition") was devised to detect systematic asymmetries in human reasoning which affect judgment reliability. Asymmetries were hypothesized and tested in three types of relations: 1) cause-effect, 2) condition-action-effect, and 3) object-property. The results show only minor differences in accuracy between causal and diagnostic reasonings, and mixed differences in recall-proficiency for the condition-action-effect relationships. Positive evidence was obtained for asymmetries in processing the object-property relationships.

The lack of validity-differential for cause-effect relations was surprising and prompted a second set of experiments. In decision-analysis, judgments about the likelihood of a certain state of affairs given a particular set of data (diagnostic inferences) are routinely fabricated from judgments about the likelihood of that data given various states of affairs (causal inferences), and not vice versa. This study was designed to test the benefits of causal synthesis schemes by comparing the validity of causal and diagnostic judgments against

5

"ground-truth" standards (Ref. 2 "Evidential Versus Causal Inferences: A Comparison of Validity").

The results demonstrate that the validity of causal and diagnostic inferences are strikingly similar; direct diagnostic estimates of conditional probabilities were found to be as accurate as their synthetic counterparts deduced from causal judgments. The reverse is equally true. Moreover, these accuracies were found to be roughly equal for each causal category tested. Thus, if the validity of judgments produced by a given mode of reasoning is a measure of whether it matches the format of human semantic memory, then neither the causal nor the diagnostic scheme is a more universal or more natural format for encoding knowledge about common, everyday experiences.

These findings imply that one should approach the "divide and conquer" ritual with caution; not every division leads to a conquest, even when the atoms are cast in causal phrasings. Dogmatic decompositions performed at the expense of conceptual simplicity may lead to inferences of lower quality than those of direct, unaided judgments.

2.4 List of Publications, Reports and Presentations

2.4.1 Publications

1. Pearl, J. "Entropy, Information and Rational Decisions", _International Journal of Policy Analysis and Information Systems_, Vol. 3, No. 1, pp. 93-109, July 1979.

2. Burns, M., and Pearl, J. "Evidential Versus Causal Inferences: A Comparison of Validity" to be published in _Organizational Behavior and Human Judgment_, February 1982.

3. Pearl, J., Leal, A., and Saleh, J. "GODDESS: A Goal-Directed Decision Structuring System". Proceedings of the International Congress on Applied Systems Research and Cybernetics, December 12-16, 1980, Acapulco, Mexico. Also in the Proceeding of the Fourteenth Annual Hawaii International Conference on System Sciences, January 7-9, 1981, Honolulu, Hawaii. Also submitted for publication in the _IEEE Transactions on Pattern Analysis and Machine Intelligence_.

2.4.2 Reports

4. Kim, J.H. "A Graphic System for Evaluating Decision Aids", UCLA-ENG-7915, March 1979.

5. Pearl, J. "A Goal Directed Approach to Structuring Decision Problems" A Progress Report (11-78 to 9/78), April 1979.

6. Burns, M. and Pearl, J. "Experiments in Cognitive Decomposition," UCLA-ENG-CSL-7951, August 1979.

7. Saleh, J., Leal, A., Kim, J., and Pearl, J. "Progress Toward a Goal-Directed Decision Support System, UCLA-ENG-CSL-7973, October 1979.

8. Saleh, J., Leal, A., Pearl, J. "Progress Toward a Goal-Directed Decision Support System, Progress Report (1-80 to 4/80), April 1980.

9. Burns, M. and Pearl, J. "On the Value of Synthetic Judgments," UCLA-ENG-CSL-8032, June 1980.

10. Pearl, J., Leal, A., and Saleh, J. "GODDESS: A Goal-Directed Decision Structuring System, UCLA-ENG-CSL-8034, June 1980.

11. Kim, J. "A Simulation Model for Evaluating Decision Support Systems" UCLA-ENG-81-23, February 1981.

12. Leal, A. and Bendifallah, S. "GODDESS: A Goal Directed Decision Structuring System, Program Guide and User Manual," UCLA-ENG-CSL-8103, April 1981.

2.4.3 Presentations in Conferences, Seminars and Symposiums

The Paper "GODDESS: A Goal-Directed Decision Structuring System" by J. Pearl,

A. Leal and J. Saleh was presented at the following meetings:

1. Seminar at the Johannes Kepler Universitat Linz, Linz, Austria, June 10, 1979.

2. Seminar at the Austrian Society of Cybernetic Studies, Vienna, Austria, June 21, 1979.

3. Military Operations Research Society 44th Symposium, Vanderberg Air Force Base, California, December 4-6, 1979.

4. Seminar at the International Institute for Applied Systems Analysis (IIASA), Luxenburg, Austria, July 7, 1980.

5. International Congress on Applied Systems Research and Cybernatics, Acapulco, Mexico, December 12-16, 1980.

6. Fourteenth Annual Hawaii International Conference on System Science, Honolulu, Hawaii, January 7-9, 1981.

7. The 17th Conference on Bayesian Research, Los Angeles, California, February 19-20, 1981.

8. Seminars at the Computer Science Departments, Rutgers University (April 2, 1981) and Massachussettes Institute of Technology (April 3, 1981).

The Paper "Evidential Versus Causal Inferences: A Comparison of Validity" by M. Burns and J. Pearl was presented at the following meetings:

1. Seminar at the Computer Science Department, MIT, Cambridge, Massachusetts, June 30, 1980.

2. Artificial Intelligence and Simulation of Behavior (AISB) Conference, Amsterdam, The Netherlands, July 2-5, 1980.

3. The International Congress on Applied Systems Research and Cybernetics, Acapulco, Mexico, December 12-16, 1980.

4. The 17th Conference on Bayesian Research, Los Angeles, California, February 19-20, 1981.

2.5 Research Staff

The research staff engaged in this project include:

Dr. Judea Pearl - Principal Investigator

Dr. Norman Dalkey - Faculty Associate

Dr. Semyon Meerkov - Visiting Associate Research Engineer

Dr. Antonio Leal - Visiting Associate Research Engineer

Dr. Joseph Saleh - Graduate Student, Engineering (Ph.D., 1979)

Jin Kim - Graduate Student,Engineering (MSC. 1979)

Tsui Lavi - Graduate Student,Engineering

Salah Bendifallah - Graduate Student, Engineering

Dr. Michael Burns - Graduate Student, Psychology (Ph.D., 1980)

Robert Fiske - Graduate Student, Psychology

3.0 EXPERIMENTAL EVALUATION OF GOAL-DIRECTED STRUCTURING PROCEDURES

3.1 Approach

One of the claims stated in favor of goal-directed structuring methods has been their promise to induce the user to consider a richer set of options than that induced by other structuring methods. This expectation stems from the fact that goal-directed structuring induces the decision-maker to first consider goals and intentions and only then to recall options available for furthering each goal separately. The main objective of the experiments reported in this document has been to submit this claim to a systematic and controlled test.

The basic hypothesis that goal-directed procedures induce a richer set of alternatives has already been given an empirical comfirmation by Pitz et al. (Pitz, G.F., Sachs, N.J. and Heerboth, J., "Procedures for Eliciting Choices in the Analysis of Individual Decision", Organizational Behavior and Human Per-formance, Vol. 26, P. 396-408, 1980). Of several candidate procedures tested for evoking a wider variety of choices, the one based on subgoal elicitation was found to be most effective. In these experiments, however, the degree of variety exhibited by a given set of choices, as well as their degree of relevance, were determined by the experimenter using subjective assessments. Our objective has been to give these notions more quantitative tests.

The notion of richness implies both diversity and quality. A set of wild, diverse, but obviously irrelevant or ineffective actions would hardly be categorized as rich. The reason that richness is a meritorious quality stems from the hope that a diverse set of alternatives is more likely to contain those choices which can solve the problem satisfactorily, in much the same way that scattered shots are more likely to hit an unseen target than shots aimed at the wrong direction.

These considerations lead to several methods of measuring richness. An indirect measure would simply focus on quality. Hopefully, a method which induces a person to consider a wider set of options would also make it more likely for that person to select an effective action from the set, and consequently, exhibit a higher overall performance. Thus, the overall performance score achieved by a game playing subject could constitute an indirect measure of the richness of alternatives considered by that subject. One may argue, however, that people often lack the insight or computational power necessary for identifying a good alternative, even when such is brought to their attention, so richness and performance would correlate only weakly.

A more direct way of testing for richness would be to examine the entire set of choices considered by the subject, select the one with the highest merit (assuming an objective figure of merit can be assigned to each choice) and take its merit measure to signify the richness of the set considered. In cases where the choices could be represented by points in some topological space a still more direct measure of richness can be devised. One could then obtain a direct measure of diversity (ignoring quality) by computing the mean inter-point distance.

In the experiments conducted at our laboratory we devised a test bed possessing the last two features. Subjects were motivated to master the playing of a computer-simulated business game. An objective measure of action quality was computable via the loss-of-opportunity criterion (see Ref. 11). Additionally, each action consisted of assigning numerical values to four decision variables and could, therefore, be represented as a vector in a four-dimensional space. These factors enabled us to compute several measures of richness and quality and to test whether goal-directed procedures induce substantially different choice sets than those induced by other structuring methods (such as decision-tree elicitation).

11

3.2 Methods

Subjects. Students were recruited in two ways: announcements were posted in the business school and an advertisement was placed in the campus' student newspaper. As subjects signed up, they were given an orientation which consisted of verbal and written descriptions of the logistics of the experiment and of the business game, as well as a demonstration of the game itself. Each subject was paid an hourly wage for his or her participation in the study. As an added incentive for learning the subtleties of the business game, the second phase of the experiment was organized as a contest. Specifically, the rank ordering of subjects in terms of accumulated profit of their fictitious businesses at the end of the experiment determined the size of each person's bonus award. The graduated series of bonus awards to be used in the contest was shown to subjects before the beginning of their involvement in the experiment. Fifteen students were signed up as bona fide subjects with an additional eight put on a waiting list. Despite this number of people, there was substantial attrition, such that a total of ten subjects completed both phases of the experiment.

Procedure. There were two phases in the experiment, in both of which subjects played the computer business game. The first of these consisted solely of training. Subjects were instructed to learn as much as they could about accumulating profit for their fictitious businesses without regard to the need to avoid errors. To assist them at this, the computer was programmed to provide them with the option of starting over (i.e., returning to the initial state at time period zero) each time they logged on. The first phase lasted for a minimum of five paid hours, after which subjects were told that they could continue training without pay as long as they wished before beginning the second phase. It was explained to them that it was desirable for them to learn as much about the game as they could, but our limited funds prevented us from paying for the extra training.

12

For the second phase of the experiment, subjects were randomly assigned to one of three conditions (representing three kinds of questionnaires), henceforth referred to as the goal-directed (GD) condition, the tree-elicitation (TE) condition, and the control condition. During the second phase, subjects played one continuous game for forty business periods, without the option of starting over. Thus, the state of the simulated industry at the point at which each person logged off the computer was restored when he or she logged on again. The computer was also programmed to interrupt the play after the ninth, nineteenth, twentyninth and thirtyninth periods were completed. It elicited from subjects their decisions regarding the action to be taken in the subsequent period, and then instructed them to fill out a questionnaire before proceeding with the game. They were not shown the outcome of the before-questionnaire action. After completing the questionnaire, they were allowed to revise the decisions they made just prior to filling out the questionnaires, and then the play resumed based on the revised action. The experiment concluded after each subject revised his or her decisions regarding period forty of the game.

Each subject participated independently of other subjects. There was one computer terminal available, and subjects reserved its use ahead of time on a sign-up sheet. Each student was encouraged to sign-up for two one-hour sessions per week. The average duration of the entire experiment for each subject was six weeks.

Materials. The problem solving environment in this study was a computer simulation of an industry consisting of two fictitious business firms. One firm was controlled by the subject while the other was controlled by a fixed computer algorithm that was part of the simulation. Subjects played independently of one another, that is, the course of one subject's game had no effect on those of the other subjects. Subjects were told that as temporary presidents

13

they were completely in charge of their firms, and that their task was twofold:
1) to accumulate as much profit as possible and 2) to leave the firm in the best
possible condition for the return of the real president. The game is played as
a series of business periods, and always starts from the same initial state in
period zero. There are built-in pauses between periods in order to allow sub-
jects to inspect a business report and to enter their decisions regarding the
values of four decision-variables for the upcoming period. The four variables
subjects were given authority to manipulate were unit price, marketing expendi-
ture, proposed production volume and the amount of raw materials being pur-
chased for the period - after - next.

Subjects were left to their own devices for discovering the optimal strategy
for accumulating profit. The key to the optimal strategy rests in the fact that
the price set by the competing firm tends to follow the subject's price by moving
in small, discrete steps towards the price maintained by the subject's firm
during the previous period. A subject who realizes this can maneuver the
competitor's price into a region containing a critical price-level and maintain
it at that level. At this critical level, the subject's firm can draw the
maximum profit possible. Clearly, achieving the highest immediate profit would
not lead to the optimal course of action. This is because maneuvering the
competitor's price in an optimal fashion requires subjects to endure short-
term losses for the good of long-term gains. Accordingly, the two measures
devised for evaluating performance quality were -- Loss of Opportunity and the
pricing profile.

The Loss of Opportunity (LOO) associated with the selection of action a_0
at state S of the game is the difference between the overall future earnings
realizable from S by the optimal strategy, and that realizable from S by first
enacting a_0, then pursuing with an optimal strategy from the state which it

14

obtains. A LOO value is calculated and stored for any four-parameter action sequence that is implemented by the subjects.

The pricing profile was devised as an indicator of the quality of each subject's long-range strategy, specifically his or her understanding of the pricing relationships implicit in such a strategy. The pricing profile was administered to subjects in all conditions as the last item in every question-naire. It consisted of graph paper on which each axis was labelled in price units of each of the two firms in the industry. Subjects were instructed to plot the price their firms would establish in response to every possible price established by the competitor. For example, the diagram below depicts a pricing profile designed to maneuver the competitor's price to the neighborhood of 32, and maintain it at that level. This happens to represent the optimal strategy; however,every curve on this diagram would represent an encoding of some well-defined strategy and can be assigned a figure of merit simply by monitoring the earnings accumulated by the associated strategy over a 40-period game.
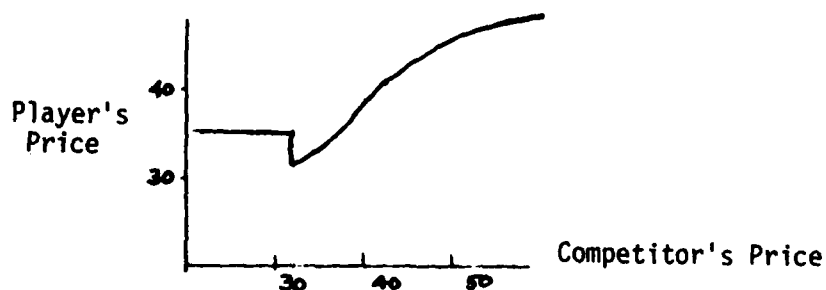
Subjects were randomly assigned to one of three conditions for the duration of the second phase of the experiment. The intervention in each condition was in the form of a questionnaire reflecting the structuring procedure being tested. Two such procedures were examined. In the GD-condition, subjects filled out a questionnaire directing them to list three objectives and two actions for accomplishing each (for a total of six actions). In the TE-condi-

tion, subjects were directed to list six mutually exclusive actions they were considering, and an exhaustive, mutually exclusive list of consequences that might possibly follow from each of them. For both conditions subjects translated the verbal description of each action into the four decision variables vector that corresponded with it. From these vectors it was possible to calculate the diversity and quality of the action set elicited by each questionnaire. Diversity was measured by the mean vectorial distance of the actions in each questionnaire. Quality was measured by the LOO of the ojectively best action in the questionnaire.

Each questionnaire is also structured to elicit from subjects numerical estimates of the viability of each action they list. For each condition, two types of estimates are made. In the GD-condition subjects estimated on a scale from 0 to 10 the level of attainment of each objective given the enactment of each action, assuming an attainment level of five before such enactment. Additionally, they rated the degree of urgency they attached to each objective on a scale from 0 to 100. Subjects in the TE-condition estimated a dollar-value and the probability of occurrence of each possible consequence they listed. In both conditions, a simple rollback procedure was used to scale each action on the basis of these estimates, and the action with the highest rollback value was designated as the action recommended by the questionnaire.

To allow for the possibility that filling out questionnaires might be a contributing factor in an observed change in performance, a third group of subjects were given control questionnaires at the same phases of the game as in the other two conditions. These consisted of questions requiring open-ended written responses, such as, "What factors influence your pricing decisions?" While this questionnaire asked subjects to articulate how much they

16

understood in the business game, it did not require them to list either
actions, goals, or possible consequences; nor to operationalize their answers
in terms of business game parameters; nor to estimate decision-making
parameters. These open-ended questions were not scored.

3.3. Results

Table 1 summarizes the background information collected on the
eight men and two women who completed the experiment as well as their
final rank ordering based on LOO measure and the number of hours
spent in the experiment. The five subjects GD_1 to GD_5 were administered
the GD questionnaire. TE_1, TE_2, and TE_3 were administered the TE ques-
tionnaire, and CT_1, CT_2 constituted the control group.

A few observations are worth noting prior to discussing the results.
The levels of understanding of the game by the subjects, as noted by in-
formal discussions with the experimentors after the training session,
varied significantly; however, those who received high scores in the
training session also obtained high ranks in the competition. Students
majoring in Business/Economics were the most motivated and, indeed,
captured the first four places in the group ranking.

While the quality of individual actions for any given subjects
fluctuated widely from period to period, we had hoped that the quality
of the pricing-profile drawn by the subjects would reflect more faith-
fully their understanding of the game and their long-range planning
ability. Instead, the actual measurement turned out to be a disappoint-
ment in this regard. The subjects had difficulty translating their pre-
ferred game strategy into a pricing-profile. We found significant contra-
dictions between the strategies portrayed in the pricing-profiles and the
game strategies actually played by the subjects. It is evident that at

17

TABLE I

Subject Code	Level in School — Freshman	Sophomore	Junior	Senior	Grad	Computer Experience	Highest Math Course Taken — Geometry	Algebra	Calculus	More than Calculus	Business-Management Experiences	GPA	Major Area of Study — Engineering	Economic-Business	Math	Theater Arts	Number of Hours Spent in Experiment	Final Ranking Based on LOO Measure
GD_1				x		x				x	0	3.0			x		14	8
GD_2				x		x			x		3	3.8		x			8	3
GD_3					x				x	x	2	3.8		x			10	4
GD_4		x							x		0	3.1				x	6	6
GD_5					x	x				x	0	3.4	x				9	7
TE_1					x	x			x		0	3.7		x			10	9
TE_2					x	x				x	0	3.6		x			12	2
TE_3		x				x				x					x		9	5
CT_1					x	x			x		0	3.3		x			9	1
CT_2			x			x			x		0	2.4	x				9	10
Frequency	0	2	1	2	5	8	0	0	5	5	5	9	2	5	2	1		
Mean												3.3					9.6	
Range												1.4					8	

least some of the subjects plotted their preferred price for the current
period not,as we had intended,as a function of the competitor's last
period price, but as a function of the competitor's price at the same
period of the game. Since we were uncertain whether a given subject
was plotting a price-profile with respect to the current price of the com-
petitor or with respect to the previous price, an analysis was made
under each assumption for every plot. Neither analysis, however, ade-
quately reproduced the rank ordering of subjects that is shown in Table 1.
Some subjects also followed very intricate strategies which could not
have been captured by a single profile (although the optimal strategy
can be expressed as a single plot). For instance, one subject seemed
to follow the lower half of his profile while raising his price and the
upper when lowering it, a distinction that can not be expressed by a
single curve profile.

Although subjects were allowed upon completion of the question-
naire to revise the decisions they made prior to filling out the ques-
tionnaires, they generally showed reluctance to utilize this option.
About half of those who did choose to revise their decision downgraded
their choice. The overall effect of the questionnaires on the perform-
ance of the players can be represented by the difference $L(AFT) - L(BEF)$,
where $L(AFT)$ stands for the LOO associated with the action selected
by the subject after completing the questionnaire (i.e., the revised
action) and $L(BEF)$ is the LOO associated with the last action chosen prior
to filling out the questionnaire. In figure 1, this difference is depicted
against the game's time period. Subjects under GD-conditions are repre-
sented by triangles $\triangle$, those under TE-conditions by circles $\bigcirc$, and
control subjects are represented by crosses $+$. Note that an upgraded

Figure 1
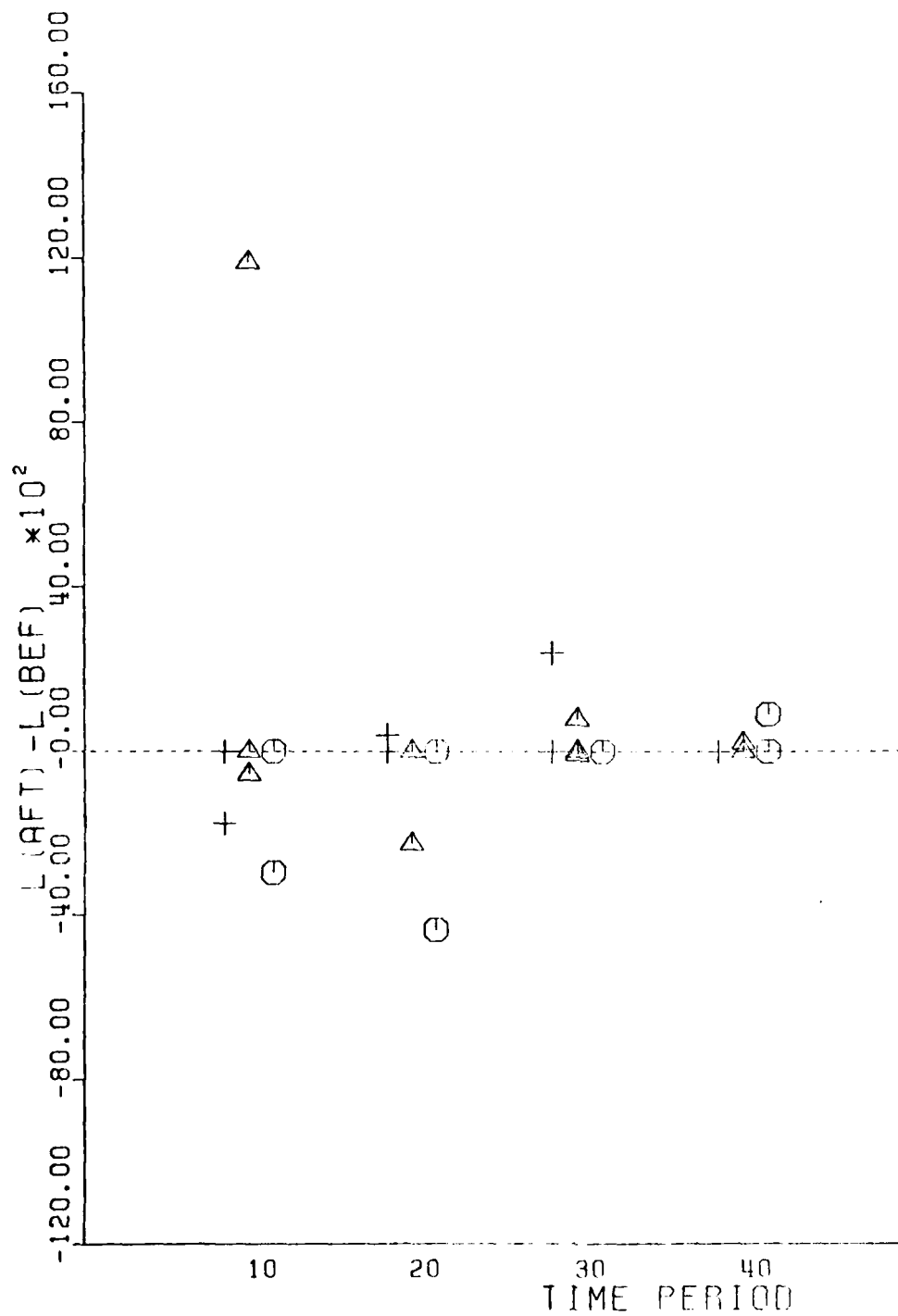
decision is represented by a negative difference L(AFT) - L(BEF). The
respective percentages of downgraded and upgraded revisions are shown
in the table below:

	GD		TE		Control	
	Cases	%	Cases	%	Cases	%
No Revisions	14	70	9	75	5	62.5
Upgraded Revisions	3	15	2	16.7	1	12.5
Downgraded Revisions	3	15	1	8.3	2	25
Total	20	100	12	100	8	100

Clearly no visible pattern emerges to distinguish any of these
groups. In order to account for the possibility that subjects were not
driven by long-range profit considerations (i.e., by the L(.) measure
which determined their monetary reward), but rather by short-range desire
to optimize the immediate profit at any given period, we also monitored
the actual immediate profit achieved by any given action. Figure 2
depicts the difference P(AFT) - P(BEF) with regard to time. Clearly, the
pattern is identical to that of Figure 1 save for the fact that a negative
difference now means downgraded revision.

The failure of subjects to properly revise their actions is not
indicative of a poor set of alternatives proposed by the questionnaire.
Indeed, Figure 3, depicting the difference L(BEST) - L(AFT), shows that
in the majority of cases subjects could have improved their performance
substantially had they possessed the insight to identify the best among
the actions which they actually considered while filling out the question-
naire. Again, no clear distinction can be detected between the GD and
TE groups.

21
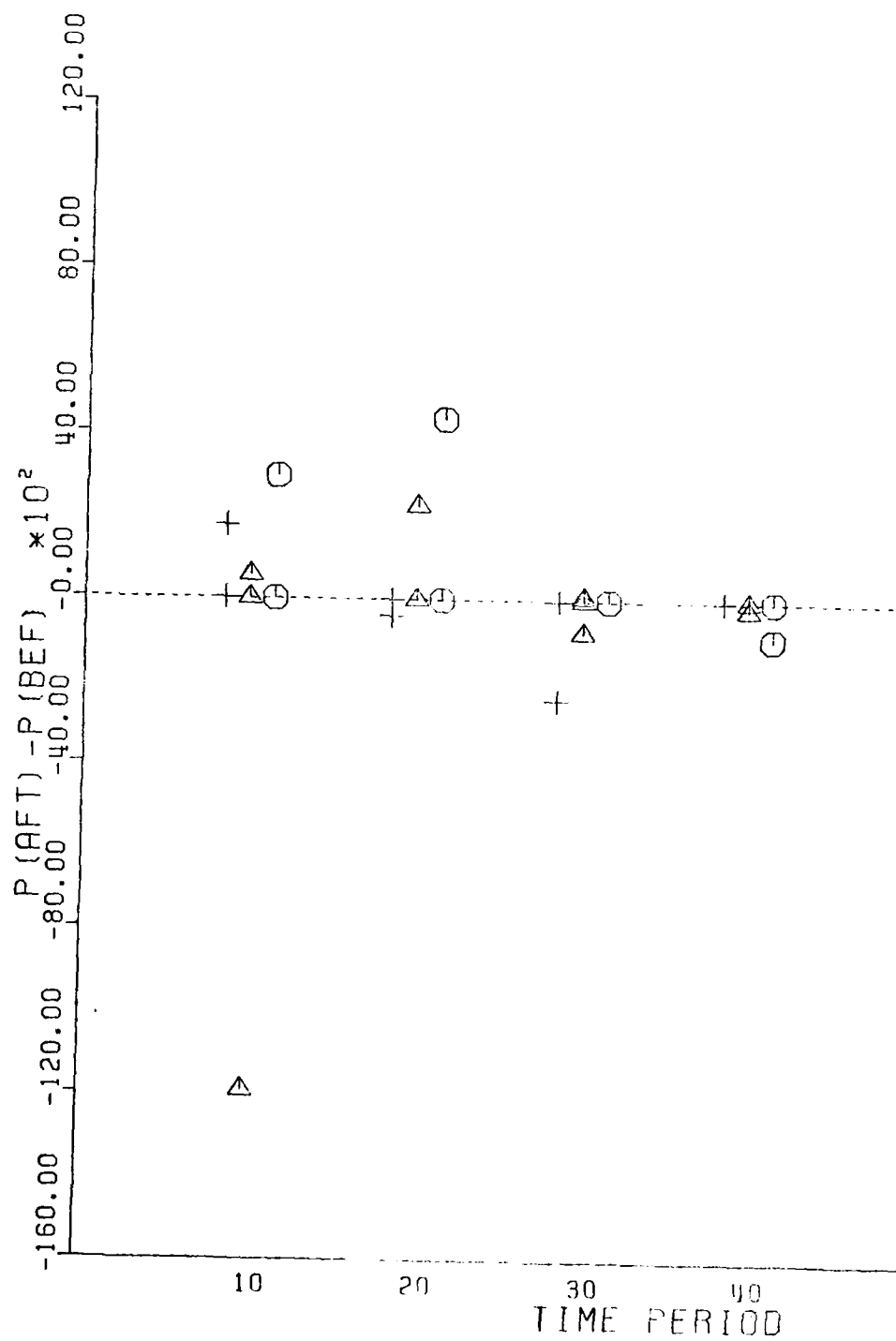
Figure 2

Figure 3

23
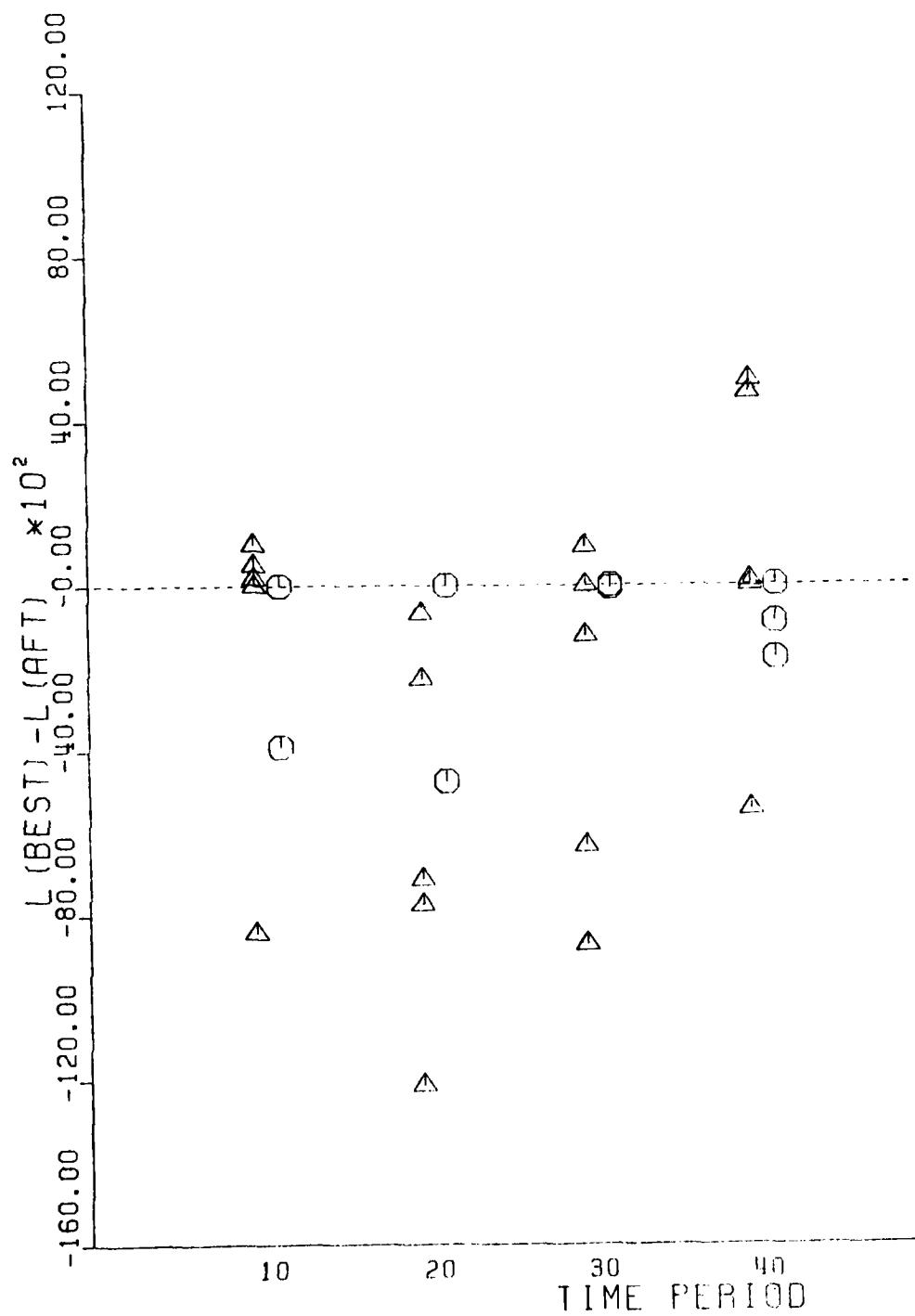
The overall effectiveness of any decision-aiding tool can be expressed as the difference between the quality of actions recommended by that tool and the quality of action selected without administering the tool. Let $L(BY)$ stand for the LOO associated with the action recommended by a questionnaire on the basis of all parameters articulated by the subject. The negative of the difference $L(BY) - L(BEF)$ would, therefore, measure the economical merit of using that questionnaire. This difference is shown in Figure 4. Seven out of the twenty actions recommended by the GD questionnaire (35%) were actually better than those originally chosen by the subjects. The corresponding figure for the TE-questionnaire is four out of twelve (33%). However, eight of the twenty actions recommended by the GD-questionnaire were worse than those originally chosen by the subjects, while only one such case occurred for the TE-questionnaire. This indicated a performance edge by the decision tree elicitation procedure. The overall means are 4635 for the GD group and -1033 for the TE group. Thus a player who is forced to comply with the recommendation derived by the decision-making tool would gain an average of 1033 units of earning potential under the TE-procedure and would lose an average of 4635 units per move under the GD-procedure. These figures represent approximately +14% and -66%, respectively, of the difference between the earnings per move generated by the optimal strategy and those generated by a typical move of the subjects.

The superiority of the TE-procedure may be attributed to the following two factors:

1) Under TE-conditions subjects were encouraged to generate and consider a more effective set of options.

Figure 4
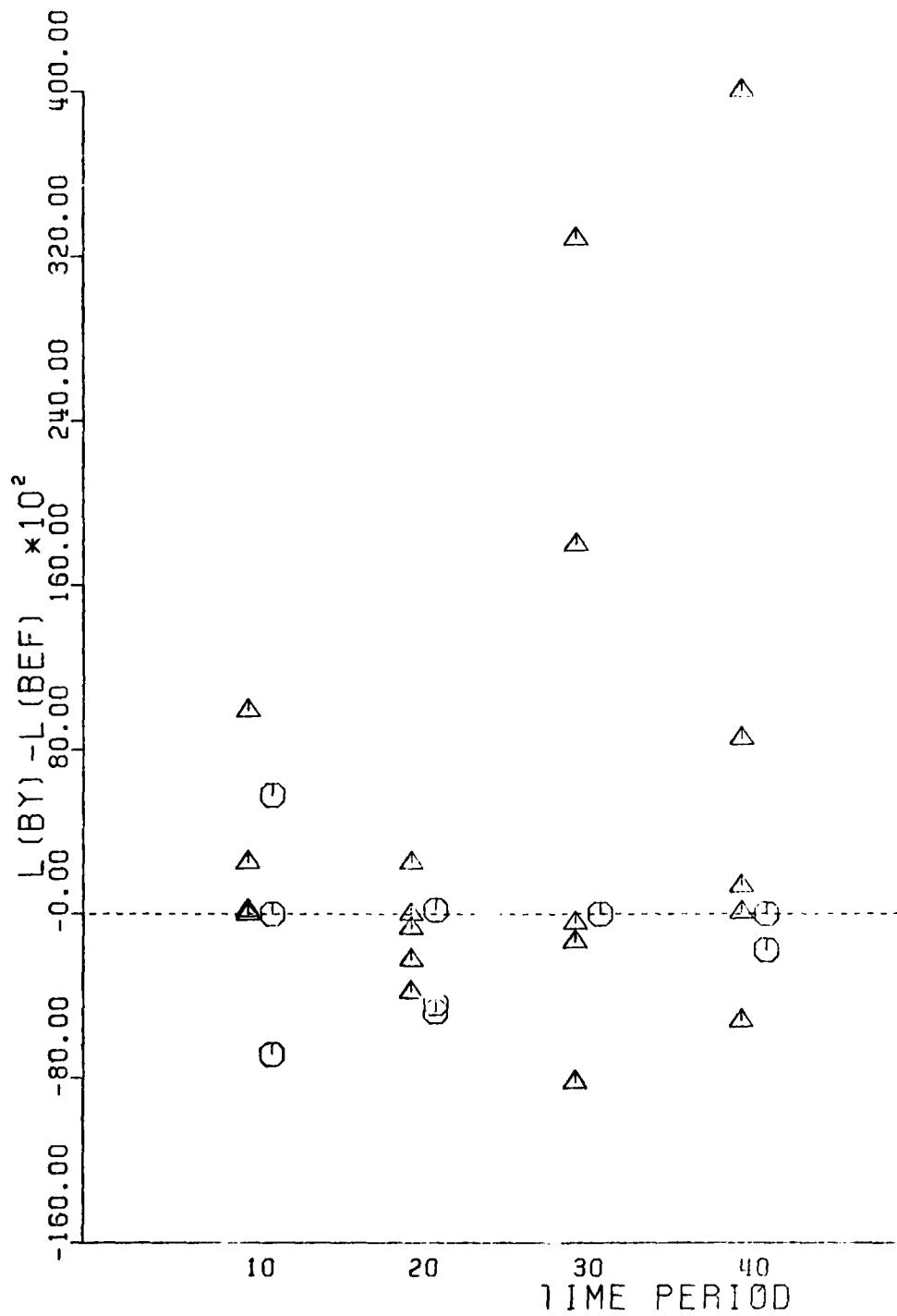
2) Under TE-conditions subjects were encouraged to articulate
a more valid set of judgments, enabling the rollback pro-
cedure to select the most promising action from the input
set of options.

An analysis of the data obtained tends to refute the first explana-
tion and support the second. Figure 5 depicts the difference L(BEST)- L(BEF)
versus the diversity of the input set of options as measured by the
(normalized) mean vectorial distance. The difference L(BEST) - L(BEF)
measures the maximum improvement in earning potential offered by a given
set of options, assuming that the subject is capable of correctly identi-
fying the best action from that set. Clearly the option sets generated
under TE-procedures (represented by circles) do not appear to contain
more effective actions than those generated under GD-procedures (represented
by triangles). On the contrary, the options generated under GD-procedures
offered an average earning improvement of 1920 units compared with the
1500 units offered by TE-procedures. In addition,the average diversity
measures of the two groups are roughly equal.

Note, however, that in only four out of twelve cases did the options
generated by TE-procedures include an action superior to that originally
played by the subjects (L(BEST) - L(BEF) < 0), as opposed to nine out of
twenty such cases for the GD-group. Moreover, in eight out of the twelve
sets produced by the TE-subjects, the original action enacted before
the questionnaire literally coincided with the best action generated
(L(BEST) - L(BEF) = 0). This happened only in four out of the twenty
sets produced by the GD-subjects.

This data supports the possibility that the two groups of subjects
utilized two distinct processes for generation. The TE-subjects apparently

Figure 5
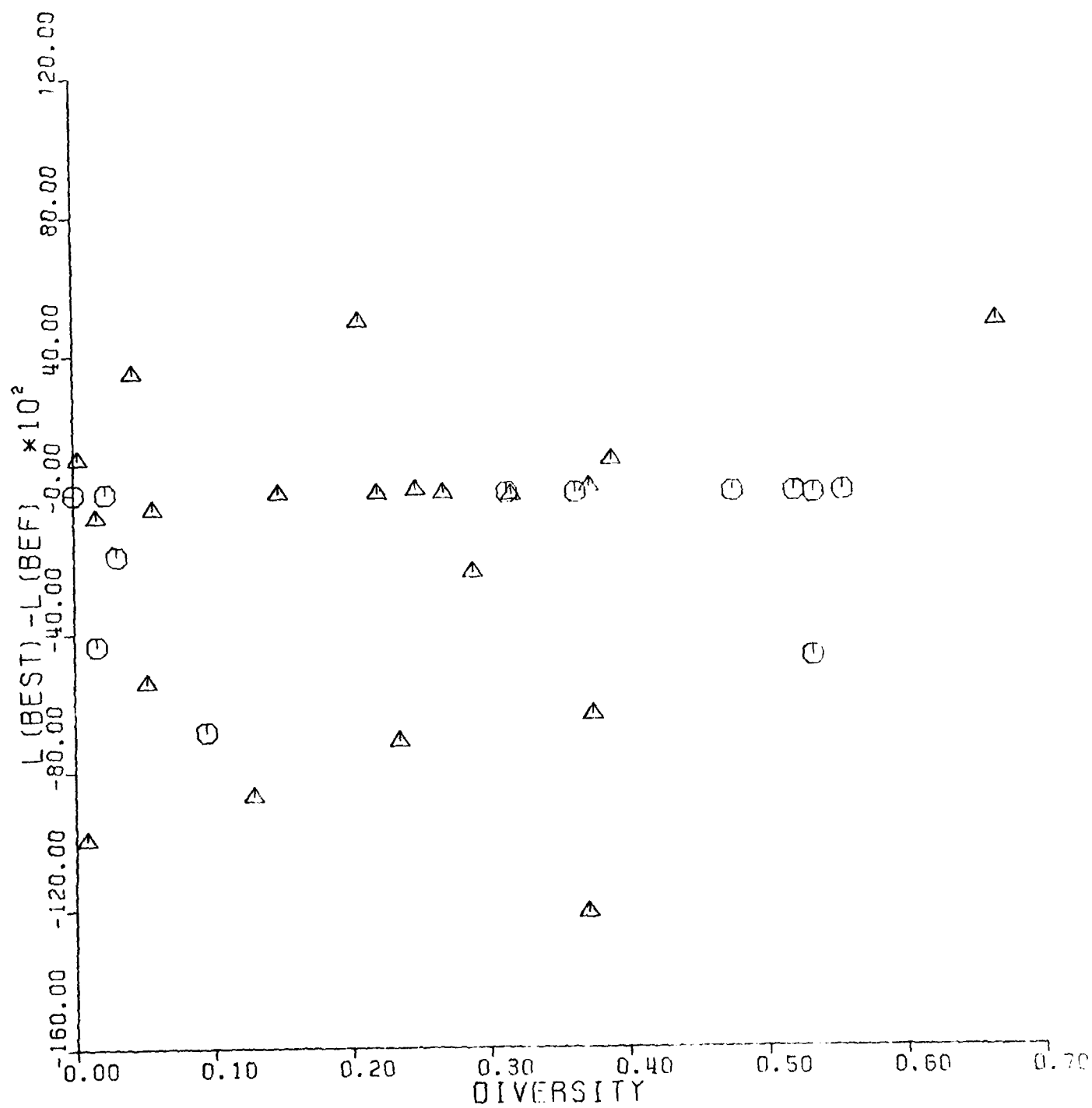
began by copying down the action just enacted and then perturbed it in various ways until the list of six required options was filled. The GD-subjects on the other hand, seemed to be generating their options afresh, with less ties to motions conceived prior to the questionnaire filling session. The GD-procedure seems to unfreeze (or deanchor) the subjects from their previous behavior. Indeed, in more than seven of twenty cases these subjects did not even list their previous actions among the set of options required by the questionnaire and fell victims therefore to the risk of generating an inferior option set. No such case was recorded among the TE-subjects.

Figure 6 explains why the TE-subjects could gain more benefit from the questionnaire if allowed to follow its recommendations. Here $L(BEST) - L(BY)$ is plotted against the diversity. The negative of this difference measures the penalty caused by the inability of the rollback procedure to identify the correct best action from a given set of options. It therefore reflects the validity (or error) of the judgment used by the players to articulate their preferences and situation assessments. Figure 6 shows that the judgments elicited by the TE-procedures were more valid than those elicited by the GD-procedures. In all but one case the TE-questionnaire successfully identified the objectively most effective action from the option sets. The GD-questionnaire selected inferior actions in more than 50% of the cases. It is significant to notice that correct identification of the best action also took place in those three cases where the previous action was inferior to the best action mentioned in the TE-questionnaires. These cases rule out the possibility that the subjects produced the option sets by a senseless perturbation around the

28

Figure 6

29
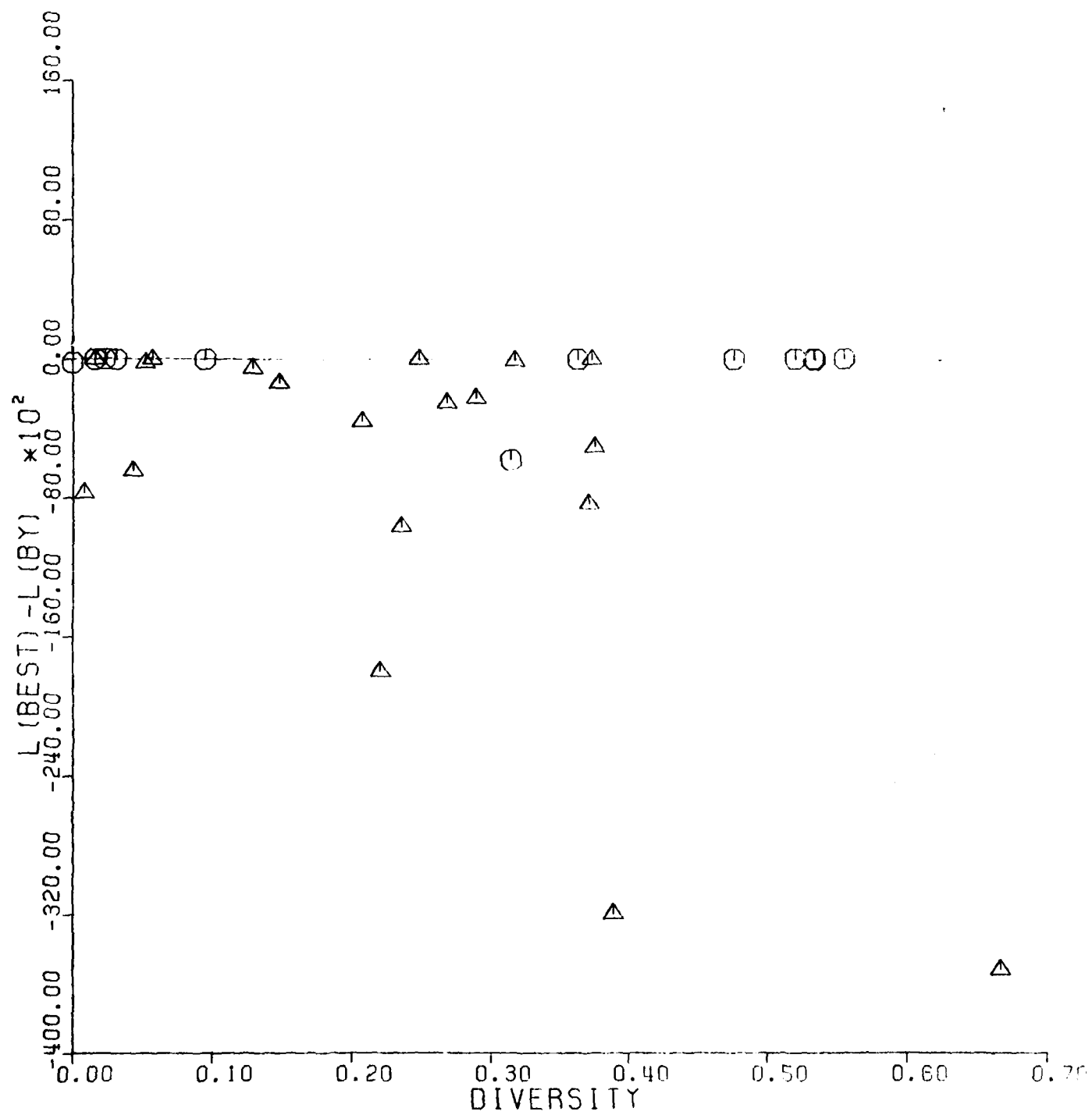
previously enacted decision, then attempted to ensure the selection of the previous decision by entering wild or overly negative judgments regarding the remaining options.

3.4 Conclusions

The experiments described in the previous paragraphs bear consequences on two different planes.

First, the methods used for evaluating the effectiveness of the two structuring procedures constitute, as far as we know, the first successful demonstration of the economical benefit associated with the use of any decision-aiding tool. The articulation of even a single level decision-tree was shown (Fig. 4) to improve the quality of decisions in a realistic, though simulated, environment. This improvement overrides the distortion which usually plagues the measurement of "objective" utility. Although the subjects were probably operating with very distorted views of the meaning of the loss-of-opportunity (LOO) measure by which they were judged, the TE-questionnaire was capable of assisting them to identify better actions than were otherwise chosen, "better" in an objective-LOO sense.

Second, our results highlight the strengths and weaknesses of the two decision-structuring methods. The goal-directed (GD) procedures exhibited superiority in setting subjects free from habitual patterns of behavior and in encouraging them to generate a novel set of options from fresh considerations. This guidance resulted in option sets which contained a higher potential for earnings improvement had the most effective action been correctly identified. The decision-tree elicitation (TE) procedure, on the other hand, permitted subjects to articulate more valid

judgments, using preference and likelihood relations, regarding the environment in which they operated. This improvement in judgment validity enabled the optimization algorithm under TE-conditions to identify the most effective action in the option set more often than the optimization algorithm under GD-conditions.

Assuming that these characteristics of the two decision-aiding processes remain the same over a wide variety of environments and planning tasks, these findings point to a method for combining the strengths of both procedures. A hybrid method utilizing the goal-directed procedure for structuring decision problems and the tree-elimination procedures for optimization would possess both merits -- the generation of novel alternatives together with a valid assessment of the environment.

We suggest, though, that the weakness in situation assessment exhibited by the GD-subjects is not characteristic of the procedure but rather that it is reflective of the unique features of the experimental environment. The goal-directed procedure and its condition-action-effect format for knowledge representation was devised to assist the structuring of long-range plans, where a long sequence of inter-related actions are to be synthesized to reach a satisfactory compromise between several objectives and requirements. None of the subjects participating in our experiments seemed to have been driven by such long-range considerations. Although the simulator was designed in such a way that the optimal strategy can only be arrived at by long-range planning sacrificing immediate profits in order to maneuver the competitor into a more desirable price range, this strategy was not discovered by any of the subjects. Instead they attempted to maximize the immediate profit, and were thus led toward reasonably profitable local maxima but prevented from realizing

31

the full earnings latent in the game. Consequently, these subjects could not exploit the full power of expression offered by the goal-directed representation; neither were they penalized by the inadquacies of the decision-tree representation in capturing complex plans. We believe that the superiority of the goal-directed approach, in both structuring and optimization, would surface in environments where the difference in performance between long-range and short-range planners is more strongly emphasized.

DISTRIBUTION LIST

CDR Paul R. Chatelier
Office of the Deputy Under Secretary
  of Defense
OUSDRE (E&LS)
Pentagon, Room 3D129
Washington, D.C. 20301

Director
Engineering Psychology Programs
Code 455
Office of Naval Research
800 North Quincy Street
Arlington, VA 22217 (5 cys)

Director
Naval Analysis Programs
Code 431
Office of Naval Research
800 North Quincy Street
Arlington, VA 22217

Director
Operations Research Programs
Code 434
Office of Naval Research
800 North Quincy Street
Arlington, VA 22217

Director
Statistics and Probability Program
Code 436
Office of Naval Research
800 North Quincy Street
Arlington, VA 22217

Director
Information Systems Program
Code 437
Office of Naval Research
800 North Quincy Street
Arlington, VA 22217

Special Assistant for Marine
  Corps Matters
Code 100M
Office of Naval Research
800 North Quincy Street
Arlington, VA 22217

Commanding Officer
ONR Eastern/Central Regional Office
ATTN: Dr. J. Lester
Building 114, Section D
666 Summer Street
Boston, MA 02210

Commanding Officer
ONR Branch Office
ATTN: Dr. C. David
536 South Clark Street
Chicago, IL 60605

Commanding Officer
ONR Western Regional Office
ATTN: Dr. E. Gloye
1030 East Green Street
Pasadena, CA 91106

Office of Naval Research
Scientific Liaison Group
American Embassy, Room A-407
APO San Francisco, CA 96503

Director
Naval Research Laboratory
Technical Information Division
Code 2627
Washington, D.C. 20375 (6 cys)

Dr. Robert G. Smith
Office of the Chief of Naval
  Operations, OP987H
Personnel Logistics Plans
Washington, D.C. 20350

Naval Training Equipment Center
ATTN: Technical Library
Orlando, FL 32813

Human Factors Department
Code N215
Naval Training Equipment Center
Orlando, FL 32813

Dr. Alfred F. Smode
Training Analysis and Evaluation
  Group
Naval Training Equipment Center
Code N-00T
Orlando, FL 32813

Mr. Milon Essoglou
Naval Facilities Engineering Command
R&D Plans and Programs
Code 03T
Hoffman Building II
Alexandria, VA 22332

CDR Robert Biersner
Naval Medical R&D Command
Code 4
Naval Medical Center
Bethesda, MD 20014

Dr. Arthur Bachrach
Behavioral Sciences Department
Naval Medical Research Institute
Bethesda, MD 20014

CDR Thomas Berghage
Naval Health Research Center
San Diego, CA 92152

Dr. George Moeller
Human Factors Engineering Branch
Submarine Medical Research Lab
Naval Submarine Base
Groton, CT 06340

Dr. James McGrath, Code 311
Naval Personnel Research and
   Development Center
San Diego, CA 92152

Navy Personnel Research and
   Development Center
Management Support Department
Code 210
San Diego, CA 92152

CDR P. M. Curran
Code 604
Human Factors Engineering Division
Naval Air Development Center
Warminster, PA 18974

Mr. Ronald A. Erickson
Human Factors Branch
Code 3194
Naval Weapons Center
China Lake, CA 93555

Human Factors Engineering Branch
Code 1226
Pacific Missile Test Center
Point Mugu, CA 93042

Dean of the Academic Departments
U.S. Naval Academy
Annapolis, MD 21402

Dr. Gary Poock
Operations Research Department
Naval Postgraduate School
Monterey, CA 93940

Dean of Research Administration
Naval Postgraduate School
Monterey, CA 93940

Mr. Warren Lewis
Human Engineering Branch
Code 8231
Naval Ocean Systems Center
San Diego, CA 92152

Dr. A. L. Slafkosky
Scientific Advisor
Commandant of the Marine Corps
Code RD-1
Washington, D.C. 20380

Mr. Arnold Rubinstein
Naval Material Command
NAVMAT 08D22
Washington, D.C. 20360

Commander
Naval Air Systems Command
Human Factors Programs
NAVAIR 340F
Washington, D.C. 20361

Mr. Phillip Andrews
Naval Sea Systems Command
NAVSEA 0341
Washington, D.C. 20362

Commander
Naval Electronics Systems Command
Human Factors Engineering Branch
Code 4701
Washington, D.C. 20360

Human Factor Engineering Branch
Naval Ship Research and Development
   Center, Annapolis Division
Annapolis, MD 21402

LDCR W. Moroney
Code 55MP
Naval Postgraduate School
Monterey, CA 93940

Professor Douglas E. Hunter
Defense Intelligence School
Washington, D.C. 20374

Dr. Robert R. Mackie
Human Factors Research, Inc.
5775 Dawson Avenue
Goleta, CA 93017

Dr. Gary McClelland
Institute of Behavioral Sciences
University of Colorado
Boulder, CO 80309

Human Resources Research Office
300 N. Washington Street
Alexandria, VA 22314

Dr. Miley Merkhofer
Stanford Research Institute
Decision Analysis Group
Menlo Park, CA 94025

Dr. Jesse Orlansky
Institute for Defense Analyses
400 Army-Navy Drive
Arlington, VA 22202

Professor Howard Raiffa
Graduate School of Business
  Administration
Harvard University
Soldiers Field Road
Boston, MA 02163

Dr. Arthur I. Siegel
Applied Psychological Services, Inc.
404 East Lancaster Street
Wayne, PA 19087

Dr. Paul Slovic
Decision Research
1201 Oak Street
Eugene, OR 97401

Dr. Amos Tversky
Department of Psychology
Stanford University
Stanford, CA 94305

Dr. Gershon Weltman
Perceptronics, Inc.
6271 Variel Avenue
Woodland Hills, CA 91364

Dr. Robert Williges
Human Factors Laboratory
Virginia Polytechnical Institute
  and State University
130 Whittemore Hall
Blacksburg, VA 24061

Dr. Meredith P. Crawford
American Psychological Association
Office of Educational Affairs
1200 17th Street, N.W.
Washington, D.C. 20036

Dr. Ward Edwards
Director, Social Science Research
  Institute
University of Southern California
Los Angeles, CA 90007

Dr. Charles Gettys
Department of Psychology
University of Oklahoma
455 West Lindsey
Norman, OK 73069

Dr. Kenneth Hammond
Institute of Behavioral Science
University of Colorado
Room 201
Boulder, CO 80309

Dr. William Howell
Department of Psychology
Rice University
Houston, TX 77001

Journal Supplement Abstract Service
American Psychological Association
1200 17th Street, N.W.
Washington, D.C. 20036 (3 cys)

Dr. Edward R. Jones
Chief, Human Factors Engineering
McDonnell-Douglas Astronautics
  Company
St. Louis Division
Box 516
St. Louis, MO 63166

Mr. Richard J. Heuer, Jr.
27585 Via Sereno
Carmel, CA 93923

Dr. John Payne
Duke University
Graduate School of Business
  Administration
Durham, NC 27706

Dr. Baruch Fischhoff
Decision Research
1201 Oak Street
Eugene, OR 97401

Dr. Andrew P. Sage
University of Virginia
School of Engineering and Applied
  Science
Charlottesville, VA 22901

Dr. Lola Lopes
Department of Psychology
University of Wisconsin
Madison, WI 53706

DATE
ILMED

-8